

## Internet Archive Blogs

*A blog from the team at archive.org*

---

# How Archive.org items are structured

Posted on [March 31, 2011](#) by [internetarchive](#)

## What is an item?

An item is a logical “thing” that we present on one web page on archive.org. An item may be one video file along with scans of the DVD cover, one book, one audio file, or a set of audio files that represent a CD , etc.

How do you know whether your files should be in one item or separate items? You get one metadata file per item. If the same metadata describes ALL of the files (like a CD), then that’s one item. If the files are too different to have the same metadata (title, creator, description, etc.), they should be in different items.

## How Items Are Structured

All archive.org items have this format URL:

**[http://archive.org/details/\[identifier\]](http://archive.org/details/[identifier])**

(where [identifier] is unique within our system).

*Example: For this item*

[http://www.archive.org/details/popeye\\_taxi-turvey](http://www.archive.org/details/popeye_taxi-turvey)

*the identifier is **popeye\_taxi-turvey***

An item is just a directory or folder of files that includes the originally uploaded content file(s) – audio, video, text, etc. – along with any derivative files we create from the originals and the metadata that describes the item. To see all files in an item, click the HTTP link in the upper left box on the item page (circled in red below).

INTERNET ARCHIVE


Web **Moving Images** Texts Audio Software Patron Info About IA Proj

Home Animation & Cartoons | Arts & Music | Community Video | Computers & Technology | Culture & Society | News & Public Affairs | Prelinger Archives | Spirituality & Religion | Sports Videos | Videogames

Search:  Animation & Cartoons GO! Advanced Search

Moving Image Archive > Animation & Cartoons > Film Chest Vintage Cartoons > Popeye: Taxi-Turvey

**View movie**




[View thumbnails](#)  
Run time: 6:00

**Play / Download (help ?)**

|                             |           |
|-----------------------------|-----------|
| <a href="#">Ogg Video</a>   | (24.6 M)  |
| <a href="#">512Kb MPEG4</a> | (25.5 M)  |
| <a href="#">MPEG2</a>       | (137.8 M) |

[All Files: HTTP](#)

**Popeye: Taxi-Turvey (1954)**





That link takes you to a directory listing showing all original, derived, and metadata files for the item.

| Index of /14/items/popeye_taxi-turvey/         |                   |           |        |
|--|-------------------|-----------|--------|
| <hr/>  |                   |           |        |
| <a href="#">../</a>                            |                   |           |        |
| <a href="#">popeye taxi-turvey.thumbs/</a>     | 20-Nov-2008 19:42 | -         |        |
| <a href="#">popeye taxi-turvey.gif</a>         | 20-Nov-2008 19:43 |           | 161164 |
| <a href="#">popeye taxi-turvey.mpeg</a>        | 10-Mar-2005 23:12 | 144445440 |        |
| <a href="#">popeye taxi-turvey.ogv</a>         | 20-Nov-2008 20:00 | 25816690  |        |
| <a href="#">popeye taxi-turvey 512kb.mp4</a>   | 20-Nov-2008 19:55 | 26696485  |        |
| <a href="#">popeye taxi-turvey files.xml</a>   | 30-Oct-2010 05:15 | 6882      |        |
| <a href="#">popeye taxi-turvey meta.xml</a>    | 18-Jul-2006 19:24 | 1201      |        |
| <a href="#">popeye taxi-turvey reviews.xml</a> | 30-Oct-2010 05:14 | 1670      |        |
| <hr/>  |                   |           |        |

You can view information about every file in this directory by viewing the file ending in `_files.xml` (in this example, `popeye_taxi-turvey_files.xml`). Each file in the item is listed here, along with whether the source is “original” (uploaded by the user), “derivative” (derived by archive.org), or “metadata” file. You will also find a format designation, various checksums, and sometimes titles for the files.

```
- <files>
- <file name="popeye_taxi-turvey_512kb.mp4" source="derivative">
  <format>512Kb MPEG4</format>
  <original>popeye_taxi-turvey.mpeg</original>
  <md5>2856b54fdcf72d211a9f9e5605c3c11b</md5>
  <mtime>1227210907</mtime>
  <size>26696485</size>
  <crc32>bba776e6</crc32>
```

```

    <sha1>b7183aa354a026c652e2409b1da27d4683528907</sha1>
  </file>
  - <file name="popeye_taxi-turvey.mpeg" source="original">
    <format>MPEG2</format>
    <md5>8f3f6378140db2eafc9291eac2cb0219</md5>
    <mtime>1110496353</mtime>
    <size>144445440</size>
    <crc32>46b0cbad</crc32>
    <sha1>e03c9ad81ff801e07c7c3964f395d3acb698fa5f</sha1>
  </file>
  - <file name="popeye_taxi-turvey_reviews.xml" source="metadata">
    <mtime>1288415696</mtime>
    <size>1670</size>
    <md5>f7295a5122da83ca59b2120d9a7208ed</md5>
    <crc32>f485688d</crc32>
    <sha1>f86455ed4ddcf2061eeec8fe7ead3f2dc1c71bb3</sha1>
    <format>Metadata</format>
  </file>
  - <file name="popeye_taxi-turvey.oif" source="derivative">

```

To see all of the metadata for the item, view the file ending in `_meta.xml` (in this example, `popeye_taxi-turvey_meta.xml`). This file should list all of the pertinent information about the item, such as title, creator, description, etc. IA's metadata schema is based on Dublin Core, but it is extremely flexible. You can add any key=value pair to this file and we will store it and make it searchable in the IA search engine. (However, it may not automatically show up on the item page.)

```

- <metadata>
  <mediatype>movies</mediatype>
  <identifier>popeye_taxi-turvey</identifier>
  <publicdate>2005-03-10 16:36:11</publicdate>
  - <publisher>
    Associated Artists Productions & Famous Studios Productions
  </publisher>

```

```
- <description>
  Popeye and Bluto both run a taxi service. Bluto bullies Popeye and gets
  Animation by Tom Johnson and Frank Endres. Music by Winston Sha
</description>
<date>1954</date>
<collection>classic_cartoons</collection>
<title>Popeye: Taxi-Turvy</title>
```

Reviews, if there are any, are contained in the `_reviews.xml` file.

One thing to note: Many “display” characteristics on archive.org, among other things, work better if your item’s identifier matches your file name. So if you’re uploading a file called `popeye_taxi-turvey.mpg`, it’s best to use the identifier `popeye_taxi-turvey` (just remove the file extension). If you’re using the upload button on archive.org, put your desired identifier in the Title field of the upload form. We turn that into the identifier automatically, and then after upload you can go back into the item and change the title to something more readable.

## Archival URLs

An item’s “details” page will always be available at

[http://archive.org/details/\[identifier\]](http://archive.org/details/[identifier])

The item directory is always available at

[http://archive.org/download/\[identifier\]](http://archive.org/download/[identifier])

A particular file can always be downloaded from

[http://archive.org/download/\[identifier\]/\[filename\]](http://archive.org/download/[identifier]/[filename])

Please Note: Archival URLs may redirect to an actual server that contains the content. For example

[http://www.archive.org/download/popeye\\_taxi-turvey](http://www.archive.org/download/popeye_taxi-turvey)

currently redirects to

[http://ia600204.us.archive.org/14/items/popeye\\_taxi-turvey/](http://ia600204.us.archive.org/14/items/popeye_taxi-turvey/)

DO NOT LINK to any archive.org URL that begins with numbers like this. This refers to the particular machine that we're serving the file from right now, but we move items to new servers all the time. If you link to this sort of URL, instead of the archival URL, your link WILL break at some point.

This entry was posted in [Technical](#). Bookmark the [permalink](#).

## 5 Responses to *How Archive.org items are structured*



**Lars Aronsson** says:

April 18, 2011 at 4:40 pm

It's sad that the Internet Archive doesn't provide any structure for series or collections of items. Defining such a structure is a lot of work, and it's sad that it can't be shared with other visitors of the Internet Archive.

One example is my identification of the series and volumes of the Transactions of the Swedish Academy, which now resides on Wikimedia Commons instead of the Internet Archive, [http://commons.wikimedia.org/wiki/Category\\_talk:Svenska\\_Akademiens\\_handlingar](http://commons.wikimedia.org/wiki/Category_talk:Svenska_Akademiens_handlingar)



[Reply](#)



**internetarchive** says:

May 3, 2011 at 11:55 pm

Hi Lars,

Archive.org currently has more than 8,000 collections of items. Here is an example: <http://www.archive.org/details/harvardclassics>

If you are interested in having a collection on archive.org, you can contact us for further information at [info at archive dot org](mailto:info@archive.org).

Alexis

[Reply](#)



**[Andreas K. Förster](#)** says:

April 19, 2011 at 2:41 pm

Feature request: There should be a way to link items with each other. For example alternative recordings, or book scans with the LibriVox audio-book and so on.

The linked item should automatically be linked back.

I hope this was the right place to post such ideas...

[Reply](#)



**internetarchive** says:

May 4, 2011 at 12:29 am

Hi Andreas,

I agree, that's a great idea! We have a way to link collection pages to other related collections, but we currently don't have a way to link items to one another (other than the user including an html link in their description, of course). As our collections grow, I think having this feature will become increasingly important – even just being able to link all the different editions/languages of a book together would be nice. I'll make sure the team is aware of this request, though I can't make any promises about delivery.

Alexis

[Reply](#)

---

Pingback: [\*Downloading in bulk using wget | Internet Archive Blogs\*](#)

---

---

**Internet Archive Blogs**

*Proudly powered by WordPress.*